REDACTED

REDACTED

A STRATEGY FOR THE LABELING AND SELECTIVE ANALYSIS OF CYSTEINE, TRYPTOPHAN, METHIONINE, HISTIDINE, AND TYROSINE CONTAINING PEPTIDE FRAGMENTS AS A ROUTE TO THE ANALYSIS OF COMPLEX PROTEIN MIXTURES.

INTRODUCTION.

At various times I (FER) have discussed with people in the lab the concept of signature peptides as a route to protein identification in complex mixtures. The premise in this strategy is that most proteins have a unique amino acid sequence that is a signature for the protein. Based on the fact that liquid chromatography, capillary electrophoresis, and mass spectrometry systems are much more adept at the analysis of peptides than the intact proteins from which they are derived, the idea is that it might be easier to target ***signature peptide fragments*** of proteins for analysis than the proteins themselves. When analyzing a complex mixture of proteins the strategy would be to cleave the protein with a proteolytic enzyme, say trypsin, and then search the DNA data base for trypsin fragments that match the mass of the signature peptide and it's sisters (or are they brothers). This technique is currently used by the mass spec groups throughout the world.

The problem with this approach is that in complex mixtures containing thousands of proteins it is probable that a hundred thousand or more peptides will be generated during proteolysis. This is beyond the resolving power of liquid chromatography and mass spectrometry systems. Perhaps very high resolution multidimensional chromatographic systems coupled in tandem with MALDI mass spectrometry could handle mixtures of this complexity, but it would be very time consuming. An alternative strategy is disclosed here.

SELECTING PEPTIDE FRAGMENTS THAT CONTAIN SPECIFIC AMINO ACIDS.

This alternative strategy addresses the complexity problem while at the same time aiding in the identification of peptides selected from the mixture; say for example cysteine containing peptides. If it were possible to select tryptic fragments that contain cysteine you would greatly simplify the peptide digest mixture while at the same time you would know that i) the peptide has a C-terminal lysine or arginine, ii) the peptide has one or more arginines and lysines, and iii) the peptide has one or more cysteines. It will be shown below that a similar strategy could be used for tyrosine, histidine, and perhaps even methionine containing proteolytic fragments.

The issue is how to select proteolytic cleave fragments that contain specific amino acids. It is a common strategy in proteolysis of proteins to reduce and alkylate the sulfhydryl groups of the protein before proteolysis. Alkylation is generally based on two kinds of reactions. One is to alkylate with a reagent such as iodoacetic acid (Figure 1) or

APPENDIX
A

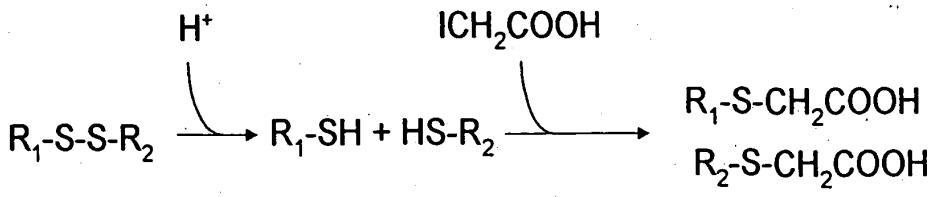


Figure 1. Disulfide reduction and alkylation with iodoacetic acid.

iodoacetamide. The other is to react with vinyl pyridine, maleic acid, or N-ethylmaleimide (Figure 2). This second derivatization method is based on the well

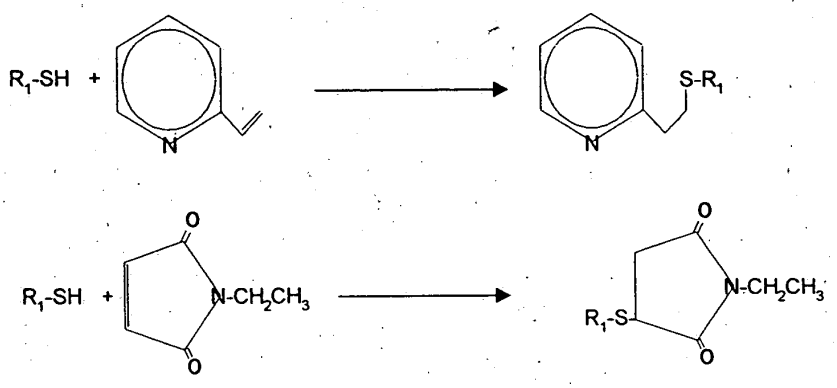


Figure 2. Alkylation of sulfhydryls by addition to conjugated double bonds.

known propensity of –SH groups to add to double bonds in a conjugated system. If the alkylating agent were to contain some sort of affinity ligand (Figure 3), it would be possible to recapture the affinity tag after the reaction and concomitantly the peptide fragment it alklyated. In this manner it would be possible to select only cysteine fragments from a mixture. Alkylation before reduction would allow one to capture only those fragments in which the cysteine was free in the native protein. Free sulfhydryl groups are even more rare.

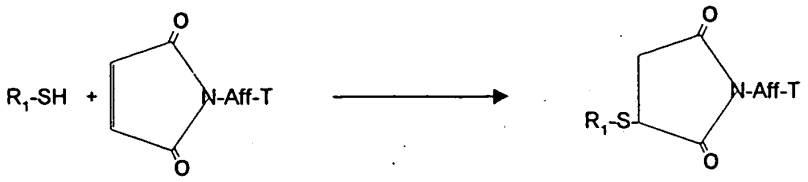


Figure 3. Alkylation of a sulfhydryl group with a maleimide affinity tag.

AFFINITY TAGGING AGENTS.

The sulfhydryl affinity tag is generated by adding it to a species that readily reacts with a sulfhydryl group. It has been noted above that –SH adds readily to maleate. Starting with an affinity tag (Aff-T) that contains a primary amine group it is possible to form the N-maleimide derivative of the affinity tag as shown in Figure 4. A specific case in which the affinity tag is a peptide R10-R12 is seen in Figure 5.

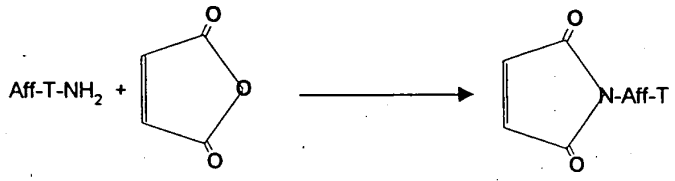


Figure 4. Formation of maleimide affinity tag from maleic anhydride.

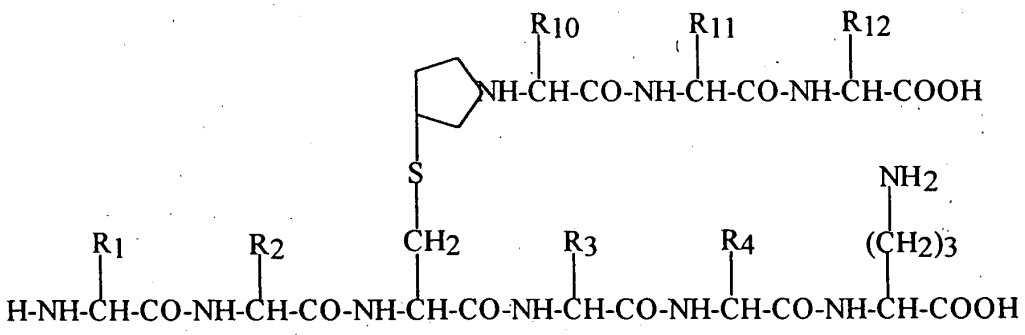


Figure 5. A cysteine peptide labeled with an affinity tag. The affinity tag is the peptide R10 - R12.

When the peptide fragments are generated by trypsin or lys-C, the tag should be a non-lysine or non-arginine containing peptide. This precludes cleavage of the affinity tag during proteolysis.

Peptide tags would most probably be subsequently captured by an immunosorbent. Polyhistidine peptides could also be used as an affinity tag. In this case they would be captured by an IMAC column. The only problem with this approach is that all other peptides in the digest that contain multiple histidine residues would also be captured. Ethlenediamine terminated biotin could be used as a tag with maleimde and captured by avidin. The negative in this approach is that avidin would capture the peptides with such great affinity that it would be difficult to release them. A short oligonucleotide or PNA sequence could also be used and be captured by hybridization.

SELECTING TYROSINE CONTAINING FRAGMENTS.

Tyrosine is another limited abundance amino acid. It is known that diazonium salts add

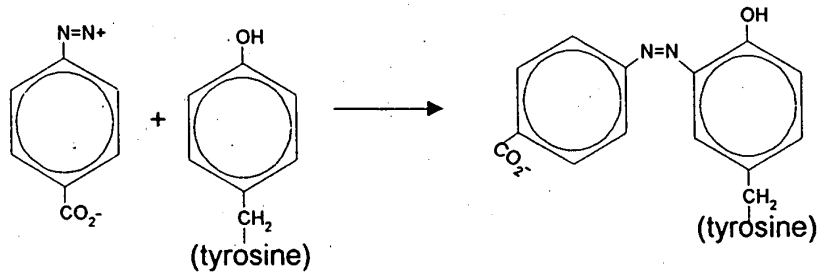


Figure 6. Reaction of tyrosine with aromatic diazonium salt.

to the aromatic ring of tyrosine *ortho* to the hydroxyl groups (Figure 6). This fact has been widely exploited in the immobilization of proteins through tyrosine. The carboxyl on the aromatic ring of the diazonium salt would be coupled to the affinity tag through a primary amine on the affinity tag as was discussed above.

SELECTING HISTIDINE CONTAINING FRAGMENTS.

Histidine rich fragments would be selected by IMAC chromatography.

ANALYSIS OF GENETIC EXPRESSION.

The objective in analyzing genetic expression is to find those proteins that are up and down regulated. It is now know that there is a poor correlation between genetic expression of mRNA, generally measured as cDNA, and the amount of protein expressed by that mRNA. Certainly mRNA concentration will change, but not necessarily in proportion to protein concentration. There are many cases where mRNA will be up regulated and protein concentration will not change at all. It would be very desirable to have a method to find only those proteins that change and then identify them.

That is now done with 2-D electrophoresis. The degree to which the concentration of a protein changes is determined by staining the gel and visually observing those spots that changed.

When it is thought that the concentration of a protein has changed, it may be quantitated with a gel scanner. Inherent in this approach is that you must also have a control 2-D gel to allow you to determine the concentration of the analyte before it was either up or down regulated.

Another strategy would be to use some type of double labeling in which the control and experimental samples are labeled with different isotopes, mixed, and analyzed simultaneously. [*Double labeling with radio isotopes has been used for a long time. It was also widely used in the past with mass spectrometry, but seems not to be used so much anymore. Heavy isotope labeling is also used to prepare internal standards.*] In this approach, the analyte is purified to homogeneity and the labeling ratio determined to know the relative concentration of analyte between the experimental and control conditions. The problem with this approach in proteins is how to label the protein. You could think in term of radio isotopes or heavy isotopes but it would require that the experiment be done with the labeled species. [It's obviously difficult to get humans to eat labeled food, let alone produce the food.] It is for this reason that post experimentation labeling is very attractive. The various labeling procedures described above allow a wide variety of isotopes to be used in a variety of post biosynthesis labeling strategies.

With the procedures described above it would be relatively easy to incorporate either N-15 or deuterium into the affinity tag used for either cysteine or tyrosine. One analytical protocol would be to label all protein in the control sample with a heavy isotope tag. All proteins from the experimental sample would be derivatized with the normal tag. Two strategies outlined below could be used for analysis. One is to cleave the protein with a proteolytic enzyme, select the tagged peptides, separate the tagged peptides in a 1-D or 2-D chromatography system and then analyze the peptides by MS. The second approach would be to separate the tagged protein first by 2-D electrophoresis and then do the proteolysis and MS analysis. In this case, selection of the tagged species either before or after electrophoresis is optional.

Theoretical mass spectra from the first approach are shown in Figure 7. . It is seen in this figure that when control and experimental samples are combined prior to analysis that peptides from down and up regulated species are easily identified. Based on the fact that there will be hundreds to thousands of peptides in a combined sample, many of them will not change in concentration between the control and experimental. These peptides will be used to establish the normalized natural/heavy isotope ratio for peptides that where neither up nor down regulated. All peptides in which the natural/heavy ratio exceed this value were up regulated. In contrast, those in which the ratio decreases were down regulated.

The beauty of this approach is that it is an internal standard method that detects relative change, not absolute amount. It is very difficult to determine relative changes in analytes that are present at very low levels. This method is as sensitive to changes in very dilute analytes as it is those that are present at great abundance. Another great advantage of this approach is that it is not influenced by quenching in the MALDI. This means that large numbers of peptides can be analyzed irrespective of the expected quenching.

PEPTIDE IDENTIFICATION.

The procedure described above allows one to scan through a complex peptide mixture from a protein digest and find those peptides that were either up or down regulated. The problem is to identify the protein from which a peptide of interest originated.

The standard protocol would be to scan the DNA data base for proteolytic fragments that also contain cysteine and match the predicted molecular weight. In many cases this will work. When it fails some other approach will be necessary. When the peptide has been separated by RPC, it will relatively pure or can be purified to homogeneity. Pure peptides can be at least partially C-terminal sequenced by MALDI. The second approach would be to sequence the peptide by MS/MS. This would be a particularly powerful approach. In the event that the peptide can not be found in any of the known DNA and protein data bases, it will be necessary to sequence the protein or the DNA from which it was derived. This can most easily be achieved by using the peptide sequence to generate a DNA sequence that is use to select the approach cDNA from a cDNA library and then DNA sequence the cDNA specific for the peptide. The cDNA thus selected could even be used to express the protein, either *in vitro* or *in vivo*.

ANALYTICAL PROTOCOL.

Strategy I. Analysis of Protein Mixtures.

- **Step 1**. Reduction of entire sample containing several thousand proteins in a robotic sample handling system.
- **Step 2**. Alklyate sulfhydryl containing peptides. When sulfhydryl selection will be done the alkylating reagent will be an affinity tagged maleimide. When the selection will be for another amino acid, the alkylating agent will probably be iodoacetic acid.
- ***Step 2'***. If another amino acid is to be affinity selected, such as tyrosine, that derivatizing agent is added at this step.
- **Step 3**. Proteolysis; generally with trypsin, but any proteolytic enzyme or combination of enzymes could be used. Enzymatic digest could either be done in the robotic system or with an immobilized enzyme column.
- **Step 4**. The experimental and isotopically labeled control samples are combined.
- **Step 5**. An affinity sorbent is used to adsorb affinity tagged species. Non-tagged peptide species are eluted to waste.
- **Step 6**. Tagged species are desorbed from the affinity sorbent.
- **Step 7**. Tagged species are chromatographically resolved. In the simplest case the sample is subjected to high resolution RPC alone. Still higher resolution can be achieved by using two dimensional chromatography. Step gradient elution ion exchange chromatography with RPC of each fraction is still probably the best choice. Given that the ion exchange column could split the tagged species into 50 fractions and the RPC column had a peak capacity of 100, it would be possible to generate 5,000 fractions for MALDI. It is estimated that the total number of sulfhydryl containing peptides would not exceed 20,000. This would mean that no sample would contain more then 2-10 peptides. MALDI should be very capable of handling 2-10 peptides per sample.
- **Step 8**. Samples are collected from the chromatographic system and transferred directly to the MALDI plates.

Strategy II. Analysis of Genetic Expression.

- . **Step 1**. Reduction of entire sample containing several thousand proteins in robotic sample handling system.
- **Step 2**. Alklyate sulfhydryl cotaining peptides from experimental sample. When sulfhydryl selection will be done the alkylating reagent will be an affinity tagged maleimide. When the selection will be for another amino acid, the alkylating agent will probably be iodoacetic acid.
- *Step 2'*. Alklyate sulfhydryl containing peptides from control sample. When sulfhydryl selection will be done the alkylating reagent will be a heavy isotope affinity tagged maleimide. When the selection will be for another amino acid, the alkylating agent will probably be heavy isotope labeled iodoacetic acid. In this way the peptide arising from the experimental sample can still be identified.
- **Step 3**. The experimental and isotopically labeled control samples are combined.
- **Step 4**. The proteins are separated by 2-D electrophoresis or 2-D chromatography. Obviously, reduction and alkylation would destroy tertiary and quaternary structure. This would have a large impact on electrophoresis and chromatography, but could still be extrapolated to the native protein sample.
- **Step 5**. Purified or partially purified proteins are subject to proteolysis; generally with trypsin, but any proteolytic enzyme or combination of enzymes could be used. Enzymatic digest could either be done in a robotic system or with an immobilized enzyme column.
- **Step 6**. Digested samples are transferred directly to the MALDI plates.

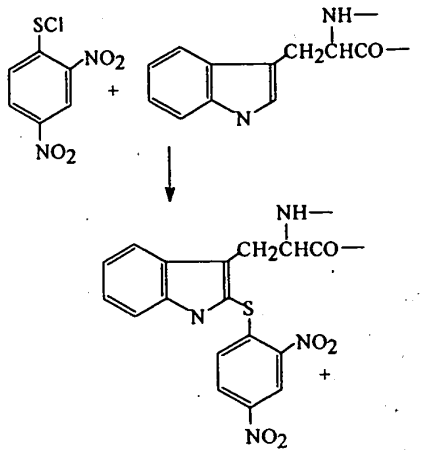


Figure 00. Derivatization of Tryptophan residue with 2,4-dinitrophenylsulfenyl chloride. [Biochim.Biophys.Acta. 278,1(1972)] See also [Can.J.Biochem.48,664(1970)][J.Chrom. 44,199(1969)][Biochem.7,971(1968)] Reaction conditions; 50% Acetic acid, 1 hr, R.T. Selection is based on DNP directed antibodies.

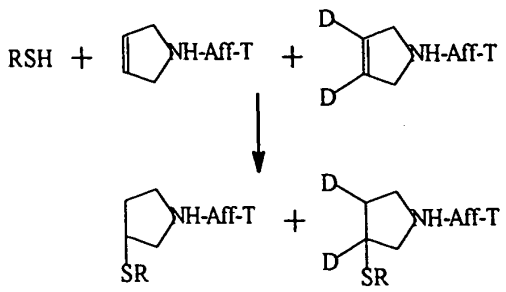


Figure 00. Derivatization of Cysteine with an affinity tagged maleimide. The function of mixing normal and deuterium labeled tag is so that tagged species are easily identified in the MALDI spectrum as a doublet that three mass units apart.

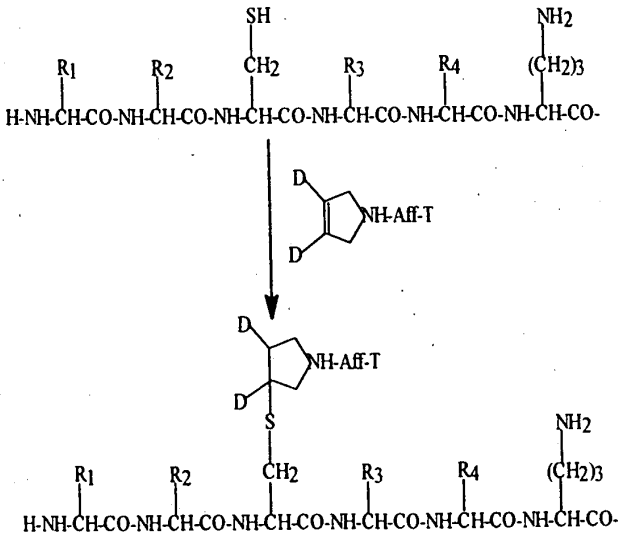


Figure 00. Derivatization of free Cysteine residue in a polypeptide with affinity tagged D_2-maleimide.

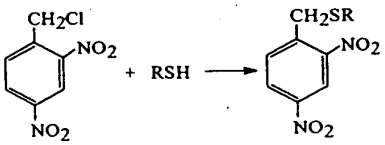


Figure 00. Derivatization of free Cysteine with 2,4-dinitrobenzyl chloride. pH 5, 1 hr. R.T.

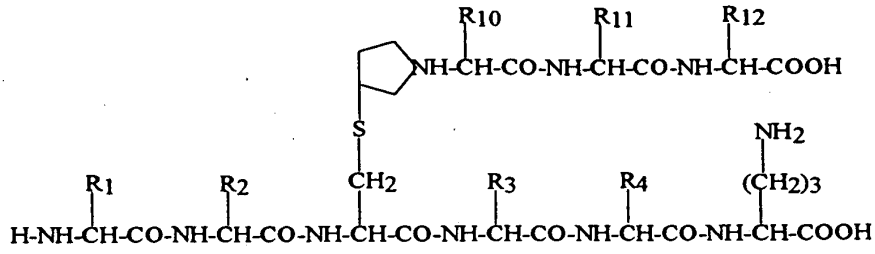


Figure 00. A polypeptide affinity tagged with the peptide R_{10}-R_{12} coupled through an N-terminal maleimide group.

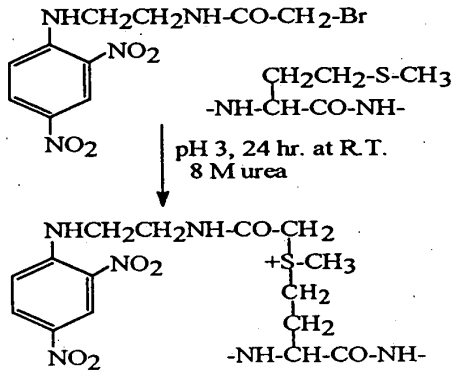
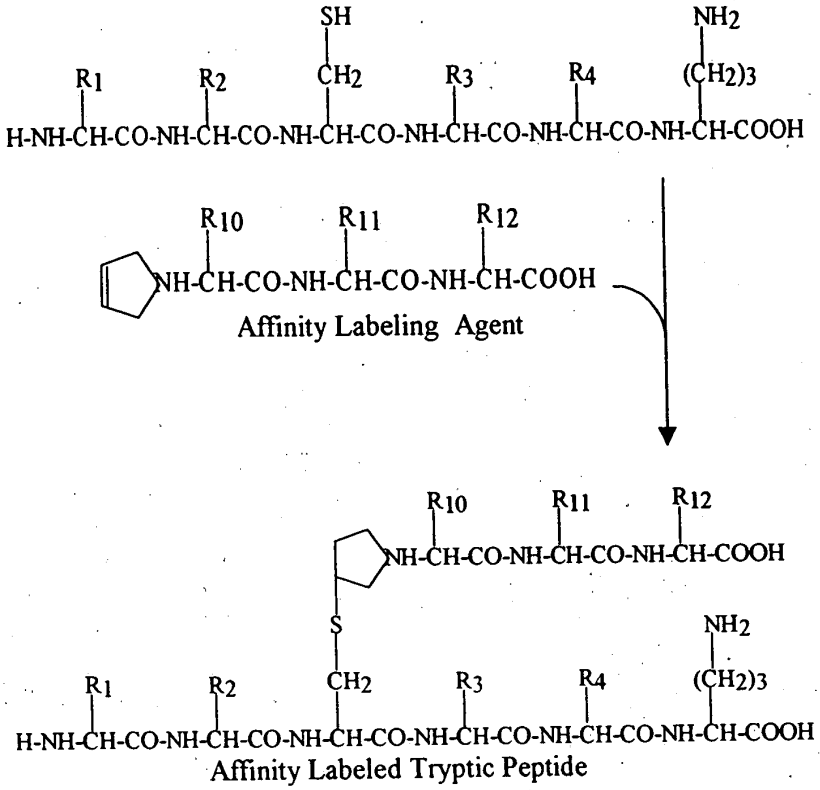


Figure 00. Dervatization of Methionine under acidic conditions. It should be noted that this derivatizing agent also derivatizes histidine at pH 5. The substantial ionization of histidine at pH 3 apparently diminishes it's alkylation. In view of the fact that histidine reacts with this reagent, it is probably best to remove histidine peptides with IMAC before derivatization.

Tryptic Peptide



Affinity Labeled Tryptic Peptide

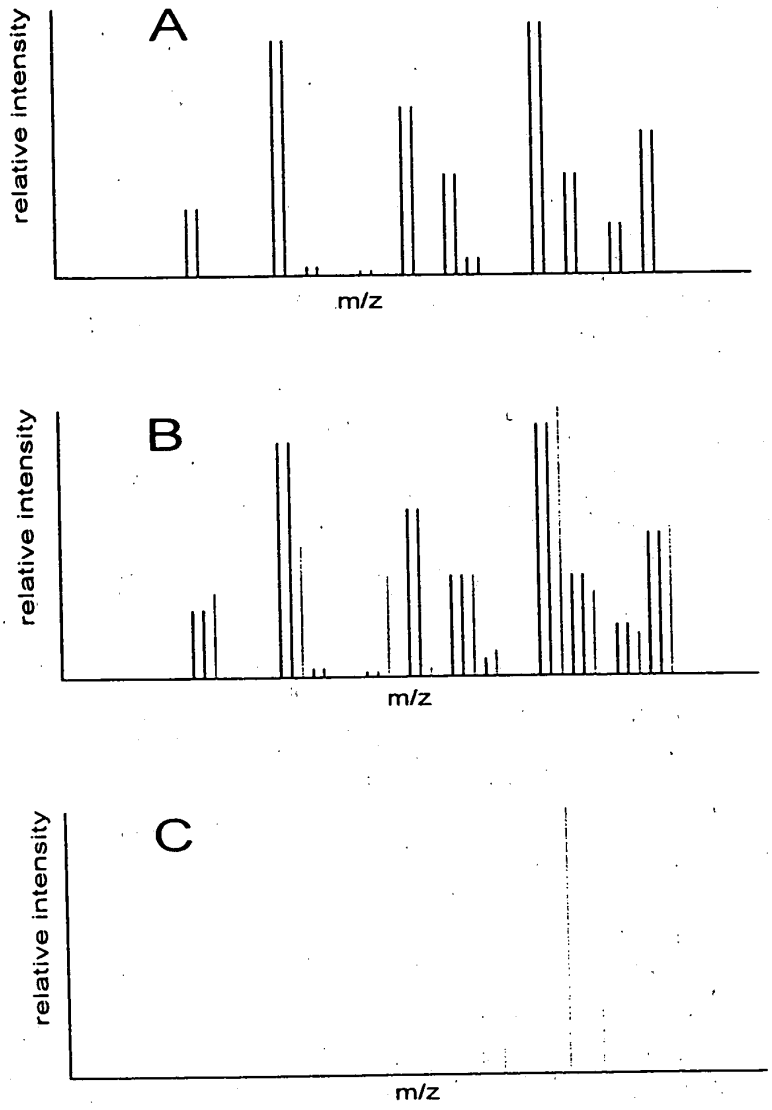


Figure 00. Heavy isotope labeling of peptides as an aid to identification and quantitation. Panel A is a mixture of native and dideutero labeled peptides. Peptides labeled with the tag are identified as doublets. Panel C is the control sample labeled with a +4 set of heavy isotopes. The +4 labeled control sample serves as an internal standard. Panel B is a mixture of the experimental and control samples. ***Notice the difficulty in identifying components that have overlapping peaks***. Also notice the problem in discriminating between control and experimental peaks because of the systematic variation of +2 and +4 labeling.

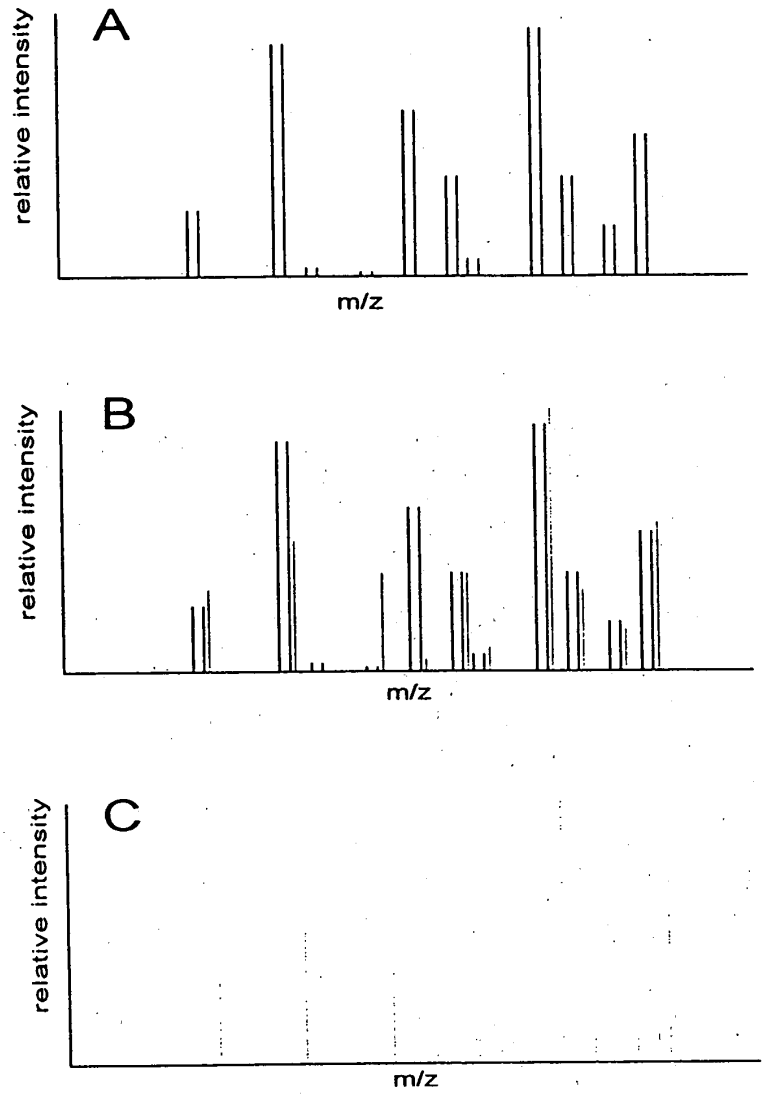
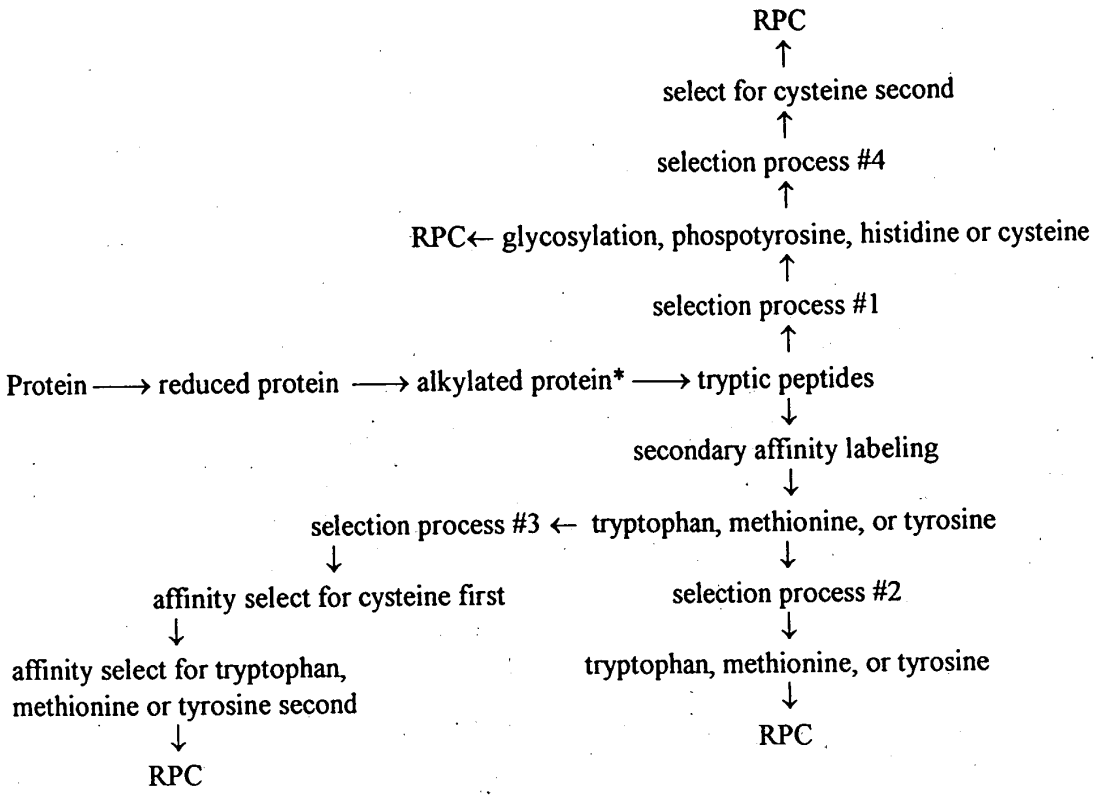


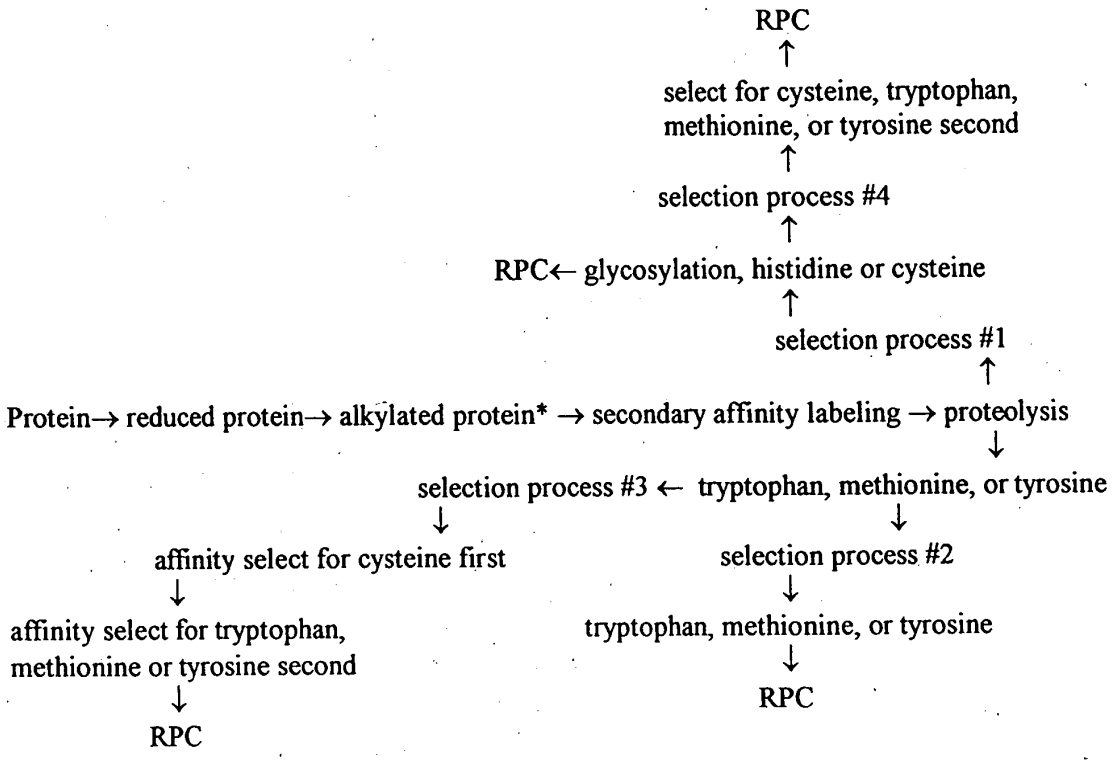
Figure 00. This Figure is identical to the above except that the contol is labeled with a +3 set of isotopes instead of the +4 as above. Notice that it is easier to discriminate between peak from the control and experimental.

Post-digestion secondary labeling protocol.



Affinity labeling cysteine residues in this case is optional. It should be noted however, that cysteine must be alkylated at this point and if it is not affinity labeled during reduction it can never be labeled.

Pre-digestion labeling protocol.



Affinity labeling cysteine residues in this case is optional. It should be noted however, that cysteine must be alkylated at this point and if it is not affinity labeled during reduction it can never be labeled.

Selective capture of specific amino acids.

1. Cysteine.
   a. Biotinylation of maleimide.
      Positives – very high affinity capture. Avidin columns are readily available.
      Negatives – it takes very acidic conditions to release from columns. A large molecule (avidin) is being used to capture a small molecule, thus a large column will be needed to get enough peptide for analysis.
   b. Histidine labeling of maleimide.
      Postivies – very simple columns may be used that are of high capacity.
      Negatives – non-cysteine containing peptides in the digest that also contain histidine will also be selected. The mass also starts to get a little high.
   c. Peptide labeling and antibody capture.
      Postives – very high capture efficiency. Easy to release captured peptide.
      Negatives - a large molecule (Ab) is being used to capture a small molecule, thus a large and expensive column will be needed to get enough peptide for analysis.
   d. Dinitrophenylation.
      Postives – very simple organic chemstry. Ab capture is very efficient.
      Negatives - a large molecule (Ab) is being used to capture a small molecule, thus a large and expensive column will be needed to get enough peptide for analysis. It is also difficult to heavy isotope label 2,4-DNP

2. Typtophan.
   a. Dintrophenylation
      Postives – very simple organic chemstry. Ab capture is very efficient.
      Negatives - a large molecule (Ab) is being used to capture a small molecule, thus a large and expensive column will be needed to get enough peptide for analysis. It is also difficult to heavy isotope label 2,4-DNP

3. Methionine.
   a. Dintrophenlyation.
      Postives – very simple organic chemstry. Ab capture is very efficient.
      Negatives - a large molecule (Ab) is being used to capture a small molecule, thus a large and expensive column will be needed to get enough peptide for analysis. It is also difficult to heavy isotope label 2,4-DNP.
   b. Histidine labeling
      Postivies – very simple columns may be used that are of high capacity.
      Negatives – non-cysteine containing peptides in the digest that also contain histidine will also be selected. The mass also starts to get a little high.
   c. Peptide labeling and antibody capture.
      Postives – very high capture efficiency. Easy to release captured peptide.
      Negatives - a large molecule (Ab) is being used to capture a small molecule, thus a large and expensive column will be needed to get enough peptide for analysis.
   d. Biotinylation.
      Positives – very high affinity capture. Avidin columns are readily available.

Negatives – it takes very acidic conditions to release from columns. A large molecule (avidin) is being used to capture a small molecule, thus a large column will be needed to get enough peptide for analysis.

4. Tyrosine.
   a. Nitrophenylation and antibody capture.
      Postives – very simple organic chemstry. Ab capture is very efficient.
      Negatives - a large molecule (Ab) is being used to capture a small molecule, thus a large and expensive column will be needed to get enough peptide for analysis. It is also difficult to heavy isotope label NP.
   b. Reaction with dianonium salts to form wide variety of derivatives.
      Postives – simple reaction that is well known.
      Negatives – very hydrophobic group, affinity tag must be attached, cross reacts with other amino acids. Could be made to work, but is down on the list of choices. Also, tyrosine is relatively abundant in proteins.

5. Histidine.
   a. Capture with an IMAC column.